Finite de Finetti theorem for conditional probability distributions describing physical theories

Matthias Christandl*
Centre for Quantum Computation, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom and Arnold Sommerfeld Center for Theoretical Physics, Faculty of Physics, Ludwig-Maximilians-Universität München, Theresienstrasse 37, 80333 Munich, Germany

Ben Toner†
School of Physics, The University of Melbourne, Victoria 3010, Australia Centrum voor Wiskunde en Informatica, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands and Institute for Quantum Information, California Institute of Technology, Pasadena CA 91125, USA

We work in a general framework where the state of a physical system is defined by its behaviour under measurement and the global state is constrained by no-signalling conditions. We show that the marginals of symmetric states in such theories can be approximated by convex combinations of independent and identical conditional probability distributions, generalizing the classical finite de Finetti theorem of Diaconis and Freedman. Our results apply to correlations obtained from quantum states even when there is no bound on the local dimension, so that known quantum de Finetti theorems cannot be used.

I. INTRODUCTION

Given a bowl containing n colored balls, we wish to compare two ways of obtaining a random sample of $k \leq n$ balls: (i) we randomly choose a ball, replace it with a ball of the same color, and repeat this step k times; (ii) we do the same but don't replace the balls. If $k \ll n$, then the probability of obtaining a particular set of k balls will be almost the same in both cases [1]. This observation has profound consequences for Bayesian statistical inference, as we now describe.

Suppose we perform an experiment k times in order to estimate some physical quantity, e.g., the probability λ that a muon decays in a given time. Let $A_i = 1$ if the ith muon decayed and $A_i = 0$ if it did not. If we assume that the results of the experiments are independent, we can posit some prior probability distribution $m(\lambda)$ and analyze our data by updating this probability distribution as more data arrives. Statisticians of de Finetti's subjective school [2] are not willing to accept this assumption, however, since for them all probability distributions should be subjective degrees of belief, which $m(\lambda)$ is not. Instead, they make the weaker assumptions that the experiment could have been performed $n \gg k$ times and that there was nothing special about the experiments actually performed. These assumptions, together with the observation about colored balls above, can be shown to imply that there exists a distribution $m(\lambda)$ such that

$$P[A_1, \ldots, A_k] \approx \int dm(\lambda) P_\lambda[A_1] \cdots P_\lambda[A_k], \tag{1}$$

i.e., the probability distribution $P[A^k]$ behaves *as if* the experiments really were independent and there really were some objective prior $m(\lambda)$. This is a statement of the famous *de Finetti representation theorem* [1, 3]. Our results establish the same correspondence for measurement results in a more general, probabilistic, physical theory, where the state of a system is described by a conditional probability distribution.

We now give a brief description of the setting and our results; precise definitions are given later on. A physical system in a probabilistic physical theory is made up of a number of—in our case identical—subsystems, called *particles*. On each particle different measurements from a set $\mathcal{X}$ can be performed and outputs from a set $\mathcal{A}$ are obtained. The state of a particle is specified by a *conditional probability distribution* $P[A|X]$: the probability of obtaining result a when performing measurement x is given by $P[A = a|X = x]$. The possible states of n particles are the conditional probability distributions $P[A^n|X^n]$ that obey a *no-signalling* property, which ensures that the reduced state on a

*Electronic address: christandl@lmu.de
†Electronic address: bentoner@bentoner.com

subset of the particles is always well-defined.

Our main result is that the joint state $P[A^k|X^k] = P[A_1 \cdots A_k|X_1 \cdots X_k]$ of k particles randomly chosen from n particles—or equivalently, the state of the first k particles of a permutation-invariant state of n particles—can be approximated by a convex combination of identical and independent conditional probability distributions,

$$P[A^k|X^k] \approx \int dm(\lambda) P_\lambda[A|X]^{\times k} \tag{2}$$

and that the error in the approximation is bounded by $|\mathcal{X}|k(k-1)/n$ in the appropriate distance measure, where $|\mathcal{X}|$ is the number of different possible measurements. (We write $P_\lambda[A|X]^{\times k}$ for $P_\lambda[A_1|X_1] \cdots P_\lambda[A_k|X_k]$.) Our result generalizes the finite de Finetti theorem of Diaconis and Freedman, who proved for classical probability distributions ($|\mathcal{X}| = 1$) that the error in the approximation is no more than $k(k-1)/n$ [1] [23].

This paper is motivated by recent work on finite *quantum* de Finetti theorems, i.e., statements of the form

$$\rho^k \approx \int d\sigma \; \sigma^{\otimes k}, \tag{3}$$

where ρ^k is the k-particle reduced density matrix of a permutation-invariant density matrix of n particles with state space of dimension d, where the error is at most $4d^2k/n$ in the trace distance [4, 5] [24]. In fact, it is necessary that the error depends on d [5], and so the quantum de Finetti is not useful in applications where d cannot be bounded. Our results are designed to apply in this setting: provided we have a bound on the number of ways $|\mathcal{X}|$ that a system is measured, the approximation in Eq. (2) will be good, even if there is no bound on the local dimension d. In recent years, quantum de Finetti theorems, especially Renner's so-called 'exponential' version [6], have been used to prove the security of quantum key distribution (QKD) schemes [7]. At the same time, attempts have been made to lift the assumption of a fixed (finite) local dimension [8]. Since quantum de Finetti theorems are necessarily dimension-dependent, they cannot be used in this setting. Although our theorems do not directly lead to security proofs either, we regard them as a first step towards this goal.

We also prove a finite quantum de Finetti theorem for separable ρ^n: in this case there is an approximation of the form in Eq. (3) with error $k(k-1)/n$, *independent* of the dimension. We do not, however, know whether our techniques can be extended to prove the finite quantum de Finetti theorem in full generality. The issue is that our theorem concerns conditional probability distributions that arise from measuring quantum states and not the quantum states themselves. If we take, for example, a tomographically complete set of measurements, the representation described in Eq. (2) will in general contain distributions $P_\lambda[A|X]$ that cannot be obtained by performing the tomographic measurements on quantum states. One can, however, apply the argument of [9] to obtain the infinite quantum de Finetti theorem and indeed an infinite de Finetti theorem for any physical theory in what is known as the *convex sets framework* [10, 11] (see [12] for the details).

Another application of our work is to the study of classical channels. Fuchs, Schack and Scudo have used the Jamiolkowski isomorphism to transfer the infinite quantum de Finetti theorem ($n = \infty$, $k < \infty$) [9, 13] to quantum channels [14]. Since a conditional probability distribution can be viewed as a classical channel with probability distributions as input and output, our results also provide a de Finetti theorem for classical channels.

Outline.—Our first task is to define an appropriate distance measure on states of k particles in probabilistic theories, in order to quantify the error in Eq. (2). The distance between states should bound the probability of distinguishing them by measurement, and so we need to be clear about what measurement strategies are allowed. One possibility, which we explore in [15], is to restrict to strategies where each of the k particles is measured individually. But when the conditional probability distributions arise from making informationally complete local measurements on entangled quantum states, the resulting distance measure fails to bound the trace distance between the quantum states. In the next section we show how to define a 'good' distance measure in which all noncontextual measurements are allowed, including all joint quantum-mechanical measurements. We then state and prove our results. In the last section, we explain the origin of the distance measure, the convex sets framework, which allows us to conclude with an open question on finite de Finetti theorems in this more general setting.

II. A DISTANCE MEASURE FOR CONDITIONAL PROBABILITY DISTRIBUTIONS

When we measure a quantum system, the probability of obtaining an outcome $a \in \mathcal{A}$ depends on which measurement $x \in \mathcal{X}$ we choose to perform on the system. It is usual to describe a quantum system using the formalism of density matrices, Hilbert spaces, and so on, but we can also describe the system by specifying a conditional probability distribution $P[A|X]$, where we write $P[A|X = x]$ for the distribution of measurement outcome A given that measurement x is performed [25]. While a classical system can be described using an *unconditional* probability distribution,

the same is not true for a quantum system, since measuring a quantum system disturbs it, eliminating our ability to make a second, incompatible, measurement on the same system.

We are therefore motivated to describe the state of an abstract system (not necessarily obeying quantum theory) using a conditional probability distribution $P[A|X]$. We view the conditional probability distribution $P[A|X]$ as the output distribution of a measurement that has been performed on system A. Alternatively, one can view $P[A|X]$ as a channel that produces an output distribution $P[A|X = x]$ on input x. For this reason we refer to the measurement setting x as the *input* and the measurement result a as the *output*. Generalizing from conditional probability distributions of one system, we shall consider a conditional probability distribution $P[A^n|X^n] = P[A_1 \cdots A_n|X_1 \cdots X_n]$, which describes an abstract system composed of n subsystems, which we call particles.

We need to be able to describe the state of a subset $\mathcal{I} \subset \{1, \ldots, n\}$ of the particles. Taking the marginal of a conditional probability distribution $P[A^n|X^n]$ yields a conditional distribution $P[A_{\mathcal{I}}|X^n]$, where the outputs at the particles in $\mathcal{I}$ depends on the inputs at all n sites. In order to trace out the particles that are not in $\mathcal{I}$ entirely, rather than just the outputs obtained from measuring them, we need another notion, that of a conditional probability distribution being *no-signalling*.

Definition 1. A conditional distribution $P[A^n|X^n]$ is *no-signalling* if for all subsets $\mathcal{I} \subset \{1, \ldots, n\}$ with complements $\bar{\mathcal{I}} := \{1, \ldots, n\} \backslash \mathcal{I}$

$$P[A_{\mathcal{I}} = a_{\mathcal{I}}|X_{\mathcal{I}} = x_{\mathcal{I}}] := \sum_{a_{\bar{\mathcal{I}}}} P[A^n = a^n|X^n = x^n] \tag{4}$$

is independent of $x_{\bar{\mathcal{I}}}$ for all $a_{\mathcal{I}}$ and all $x_{\mathcal{I}}$.

The terminology derives from the following fact: if we divide the n parties into two groups, $\mathcal{I}$ and $\bar{\mathcal{I}}$, then, provided $P[A^n|X^n]$ is no-signalling, it is impossible for the group $\mathcal{I}$ to send a signal to the group of $\bar{\mathcal{I}}$ just by changing their inputs. Not all conditional probability distributions are no-signalling; for example, $P[A_1 = a_1, A_2 = a_2|X_1 = x_1, X_2 = x_2] = [a_1 = x_2][a_2 = x_1]$ (where $[t]$ is 1 if t is true and 0 otherwise) is signalling. We note that any conditional probability distribution that arises from making local measurements on a quantum state is no-signalling. The no-signalling requirement is the minimal assumption necessary to ensure that state of any subset of particles is well-defined.

The goal of this paper is to approximate by product distributions a no-signalling conditional probability distribution on k particles arising from a symmetric conditional probability distribution on n systems, so we need to introduce a notion of distance for conditional probability distributions. This distance measure should generalize the classical variational distance, which is equal to the maximum probability of distinguishing two probability distributions, and the quantum trace distance, which is equal to the maximal probability of distinguishing two quantum states. In order to define a *trace distance* for no-signalling conditional probability distributions we therefore need to determine what measurement strategies can be used to distinguish two conditional probability distributions. In fact, there are three natural sets of measurement strategies for conditional probability distributions, each of which induces a distance measure on conditional probability distributions. We will work with the largest of these sets giving the strongest notion of a distance, for if we can show that two conditional probability distributions are almost indistinguishable using a particular set of measurements, it will trivially follow that they are also almost indistinguishable when only a subset of those measurements is allowed. Let us start by introducing the three sets.

An *individual measurement* is a distribution $P[X^k]$ on the inputs that maps the conditional probability distribution to the unconditional probability distribution $P[A^k X^k] = P[A^k|X^k]P[X^k]$. Such a measurement can be carried out by measuring each subsystem individually. Note that individual measurements also make sense if we drop the condition that $P[A^n|X^n]$ is no-signalling. Since we restrict to no-signalling conditional probability distributions, a larger class of measurements is possible and indeed needed for applications. Suppose the conditional distribution $P[A_1A_2|X_1X_2]$ is no-signalling. We start by writing

$$P[A_1A_2|X_1X_2 = x_1x_2] = P[A_1|X_1X_2 = x_1x_2]P[A_2|A_1, X_1X_2 = x_1x_2] \tag{5}$$

$$= P[A_1|X_1 = x_1]P[A_2|A_1, X_1X_2 = x_1x_2], \tag{6}$$

where we made use of the no-signalling principle, Eq. (4), in the second line. This provides an operational means to sample from $P[A_1A_2|X_1X_2 = x_1x_2]$: We first sample a_1 from the distribution $P[A_1|X_1 = x_1]$, then sample a_2 from $P[A_2|A_1 = a_1, X_1X_2 = x_1x_2]$. The important point is that a no-signalling conditional probability distribution can provide the output on system 1 before specifying which input is chosen for system 2. Therefore the following *adaptive measurement* on $P[A_1A_2|X_1X_2]$ is possible: Input x_1, obtain a_1, and choose an input $x_2 = f(a_1)$, where $f : \mathcal{A} \to \mathcal{X}$ is an arbitrary function. Such a strategy can lead to a higher probability of distinguishing two no-signalling conditional probability distributions, compared to individual strategies [26].

As in most of the paper we draw intuition from quantum-mechanical correlations. It is a well-established fact that the distinguishability of quantum states depends on whether individual or adaptive measurement strategies are considered. In the quantum case, furthermore, it is possible to apply a joint measurement to all k systems at once, a class of measurement which strictly contains adaptive measurements and can lead to strictly higher distinguishability. *Quantum data hiding* is an important application of this phenomenon [16, 17].

In defining joint operations on no-signalling conditional probability distributions, we essentially wish to allow all possible measurements whose outcomes behave like probability distributions. Motivated by this, we think of a no-signalling conditional probability distribution $P[A^k|X^k]$ as a vector in a real $|\mathcal{A}|^k|\mathcal{X}|^k$-dimensional space and consider linear functions from this space to a real $|\mathcal{A}|^k$-dimensional space. The set of *general measurements* is the set of linear functions M such that $M(P[A^k|X^k])$ is a probability distribution for all no-signalling conditional probability distributions $P[A^k|X^k]$. Clearly, individual and adaptive strategies belong to the set of general measurements, but it includes strictly more strategies, too. (The assumption of linearity is necessary so that our probability behave reasonably when we take convex combinations of states and measurements; see Ref. [18].)

Definition 2. The *trace distance* between two no-signalling conditional probability distributions $P[A^k|X^k]$ and $Q[A^k|X^k]$ is given by

$$\|P[A^k|X^k]-Q[A^k|X^k]\| := \sup_M \|M(P[A^k|X^k]) - M(Q[A^k|X^k])\|. \tag{7}$$

where the supremum is taken over all general measurements and $\|R[B] - S[B]\|$ is the classical variational distance for probability distributions $R[B]$ and $S[B]$ on system B. Extending the definition by imposing linearity, $||\cdot||$ is a norm on the space of (real) linear combinations of conditional probability distributions and hence obeys the triangle inequality.

A theory in which conditional probability distributions describe the state of a particle and where joint states of particles obey a no-signalling distribution can be treated in the *convex sets framework*. The distance measure we introduced arises naturally in this framework. We review the convex sets framework in Section IV. This will give us a broader view on de Finetti theorems and will allow us to pose an open question regarding de Finetti theorems in the convex sets framework.

III. OUR RESULTS

Suppose we have a conditional probability distribution $P[A^n|X^n]$ describing n particles. If we interchange the particles according to a permutation $\pi \in S_n$, the resulting conditional probability distribution is

$$\begin{aligned}&\pi P[A^n = a_1 \cdots a_n | X^n = x_1 \cdots x_n] \\ &= P[A^n = a_{\pi^{-1}(1)} \cdots a_{\pi^{-1}(n)} | X^n = x_{\pi^{-1}(1)} \cdots x_{\pi^{-1}(n)}].\end{aligned}$$

We say that a conditional probability distribution $P[A^n|X^n]$ is *symmetric* if it is invariant under all permutations $\pi \in S_n$. If $|\mathcal{X}| = 1$, this definition reduces to the usual definition of a symmetric probability distribution. We can now state our main result:

Theorem 3. *Suppose that $P[A^n|X^n]$ is a symmetric no-signalling conditional probability distribution. Then there exists a probability distribution p_λ such that*

$$\left\|P[A^k|X^k]-\sum_\lambda p_\lambda P_\lambda[A|X]^{\times k}\right\| \leq \min\left(\frac{2k|\mathcal{X}||\mathcal{A}|^{|\mathcal{X}|}}{n}, \frac{|\mathcal{X}|k(k-1)}{n}\right), \tag{8}$$

where the distribution p_λ is on a finite set of single-particle conditional probability distributions, labeled by λ.

This establishes that the state of a random subset of k out of n particles is well approximated by a convex combination of independent and identically distributed conditional probability distributions. To prove Theorem 3, we first show that if $P[A^n|X^n]$ is symmetric and m is chosen to be sufficiently small, then $P[A^m|X^m]$ is separable (Lemma 4). We then establish a de Finetti theorem for separable states, Lemma 5, which will complete the proof of our main result, Theorem 3. We continue with Lemma 4.

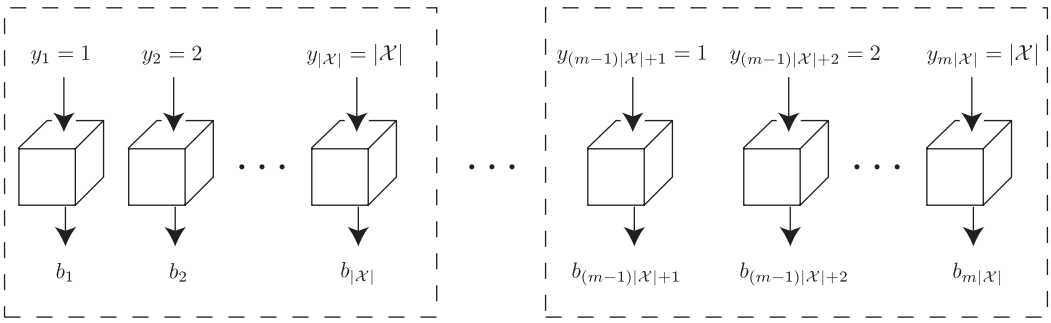

FIG. 1: Since $n = m|\mathcal{X}|$, we can divide the particles into m groups of $|\mathcal{X}|$ particles. In each of these groups we measure one particle according to each measurement in X *in advance* and record a list of all the results. In the simulation, if particle i is supposed to be measured according to a measurement $x \in X$, we just look through the ith group until we come to the particle on which measurement x was performed in advance, and output the result we find.

Lemma 4. *Let $n \geq |\mathcal{X}|$ and set $m = \lceil n/|\mathcal{X}| \rceil$. Suppose that $P[A^n|X^n]$ is a symmetric no-signalling conditional probability distribution. Then $P[A^m|X^m]$ is separable, i.e., there exists a probability distribution $p_{\lambda_1,\ldots,\lambda_m}$ such that*

$$P[A^m|X^m] = \sum_{\lambda_1,\ldots,\lambda_m} p_{\lambda_1,\ldots,\lambda_m} P_{\lambda_1}[A_1|X_1] \cdots P_{\lambda_m}[A_m|X_m],$$

where $p_{\lambda_1,\ldots,\lambda_m}$ is a probability distribution on the labels $\lambda_1, \ldots, \lambda_m$, where λ_j labels a finite set of conditional probability distributions.

Proof. In order not to obscure the main argument, we prove the statement for integral $m = n/|\mathcal{X}|$ [27]. Our technique can be traced to Werner [19]. We imagine the m particles to be separated in space and note that $P[A^m|X^m]$ is separable if and only if it can be simulated by a local hidden variable model. Such a simulation is described in Fig. 1. We now provide the formal proof. We construct a separable conditional distribution $Q[A^m|X^m]$ and then show that it is equal to $P[A^m|X^m]$. We assume that $\mathcal{X} = \{1, 2, \ldots, |\mathcal{X}|\}$, define a vector $y^n = (y_j)_{j=1,\ldots,n}$ with coordinates $y_j = (j - 1 \mod |\mathcal{X}|) + 1$, and define the separable state

$$Q[A^m|X^m] = \sum_{b^n} q_{b^n} Q_{b^n,1}[A_1|X_1] \cdots Q_{b^n,m}[A_m|X_m],$$

where $b^n \in A^n$ is distributed according to $q_{b^n} = P[A^n = b^n|X^n = y^n]$ and the single-particle conditional probability distributions are deterministic and defined by $Q_{b^n,i}[A_i = a_i|X_i = x_i] = [a_i = b_{(i-1)|\mathcal{X}|+x_i}]$, where $[t] = 1$ if t is true and 0 otherwise. Let $\mathcal{L} = \{1, 2, \ldots, n\}$, $\mathcal{L}_1 = \{(i-1)|\mathcal{X}| + x_i : i = 1, 2, \ldots, m\}$ and $\mathcal{L}_2 = \mathcal{L} \backslash \mathcal{L}_1$. Further let $A^{\mathcal{L}} = A^n$, $A^{\mathcal{L}_1} = (A_{x_1}, A_{|\mathcal{X}|+x_2}, \ldots, A_{(m-1)|\mathcal{X}|+x_m})$ and $A^{\mathcal{L}_2} = A^{\mathcal{L}} \backslash A^{\mathcal{L}_1}$ and define $b^{\mathcal{L}}, b^{\mathcal{L}_1}$ and $b^{\mathcal{L}_2}$ similarly. We find

$$\begin{aligned}
&Q[A^m = a^m|X^m = x^m] \\
&\quad = \sum_{b^n} P[A^n = b^n|X^n = y^n][a_1 = b_{x_1}] \cdots [a_m = b_{(m-1)|\mathcal{X}|+x_m}] \\
&\quad = \sum_{b^{\mathcal{L}_2}} P[A^{\mathcal{L}_1} = a^m, A^{\mathcal{L}_2} = b^{\mathcal{L}_2}|X^{\mathcal{L}_1} = x^m, X^{\mathcal{L}_2} = y^{\mathcal{L}_2}] \\
&\quad = P[A^{\mathcal{L}_1} = a^m|X^{\mathcal{L}_1} = x^m] = P[A^m = a^m|X^m = x^m],
\end{aligned}$$

where we started with the definition of $Q[A^m|X^m]$, split the summation over $\mathcal{L}_1$ and $\mathcal{L}_2$, dropped the conditioning over $X^{\mathcal{L}_2} = y^{\mathcal{L}_2}$ because of the no-signalling property of P, used the definition of a marginal state, and, lastly, the permutation-invariance of P. □

Our next statement is a de Finetti theorem for symmetric separable conditional probability distributions.

Lemma 5. *Suppose that $P[A^m|X^m]$ is a symmetric separable conditional probability distribution. Then there exists*

a probability distribution p_λ such that

$$\left\| P[A^k|X^k] - \sum_\lambda p_\lambda P_\lambda[A|X]^{\times k} \right\| \leq \min\left(\frac{2k|\mathcal{A}|^{|\mathcal{X}|}}{m}, \frac{k(k-1)}{m} \right), \tag{9}$$

where p_λ is a probability distribution on a finite set of conditional probability distributions, labeled by λ.

Proof. Let $Q_1[A|X], \ldots, Q_E[A|X]$ be the extreme points of the set of conditional probability distributions of one system. These are the deterministic functions $\mathcal{X} \mapsto \mathcal{A}$, hence $E = |\mathcal{A}|^{|\mathcal{X}|}$. Any symmetric separable conditional probability distribution is a convex combination of conditional probability distributions of the form $Q[A^m|X^m] = \frac{1}{m!} \sum_\pi Q_{i_{\pi^{-1}(1)}}[A|X] \cdots Q_{i_{\pi^{-1}(m)}}[A|X]$, where $1 \leq i_1, \ldots, i_m \leq E$. Define $Q[A|X] := \frac{1}{m} \sum_{j=1}^m Q_{i_j}[A|X]$. We expand

$$Q[A|X]^{\times k} = \sum_{j_1=1}^m \cdots \sum_{j_k=1}^m M_m(i_{j_1}, \ldots, i_{j_k}) Q_{i_{j_1}}[A_1|X_1] \cdots Q_{i_{j_k}}[A_m|X_m], \tag{10}$$

where $M_m(i_{j_1}, \ldots, i_{j_k}) = 1/m^k$ is the multinomial distribution. To compare this expression with $Q[A^k|X^k]$, write

$$Q[A^k|X^k] = \sum_{j_1=1}^m \cdots \sum_{j_k=1}^m H_m(i_{j_1}, \ldots, i_{j_k}) Q_{i_{j_1}}[A_1|X_1] \cdots Q_{i_{j_k}}[A_m|X_m], \tag{11}$$

where $H_m(i_{j_1}, \ldots, i_{j_k})$ is the hypergeometric distribution for an urn with m balls (see [1]). Then

$$\begin{aligned} \left\| Q[A^k|X^k] - Q[A|X]^{\times k} \right\| &= \Big\| \sum_{j_1, \ldots, j_k} \big(H_m(i_{j_1}, \ldots, i_{j_k}) \\ &\qquad M_m(i_{j_1}, \ldots, i_{j_k}) \big) Q_{i_{j_1}}[A_1|X_1] \cdots Q_{i_{j_k}}[A_m|X_m] \Big\| \\ &\leq \sum_{j_1, \ldots, j_k} \left| H_m(i_{j_1}, \ldots, i_{j_k}) - M_m(i_{j_1}, \ldots, i_{j_k}) \right| \\ &\leq \min\left(\frac{2kE}{m}, \frac{k(k-1)}{m} \right), \end{aligned} \tag{12}$$

where we used the triangle inequality and Diaconis and Freedman's result on estimating the hypergeometric distribution with a multinomial distribution [1]. □

These two lemmas enable the proof of Theorem 3.

Proof of Theorem 3. Set $m = \lceil n/|\mathcal{X}| \rceil$ and apply Lemma 4. Then apply Lemma 5. □

Our final result is an application to quantum theory. In complete analogy to Lemma 5 we show that the k-particle reduced state of a every separable symmetric density operator on m copies of $\mathbb{C}^d$ is approximated by a convex combination of tensor product states. Importantly, the approximation guarantee is independent of the dimension d, in contrast to the case of entangled states where a dependence on the dimension is necessary [5]. The norm is given by the trace norm $||A||_1 = \mathsf{Tr}\sqrt{A^\dagger A}$ for operators A on $\mathbb{C}^d$. It induces a distance measure on the set of quantum states that has a similar interpretation as a measure of distinguishability as the variational distance for probability distributions and the trace distance introduced on conditional probability distributions.

Theorem 6. *If ρ is a separable permutation-invariant density operator on $(\mathbb{C}^d)^{\otimes n}$, then there is a measure $m(\sigma)$ on states σ on $\mathbb{C}^d$ such that*

$$\left\| \rho^k - \int dm(\sigma)\, \sigma^{\otimes k} \right\|_1 \leq 2\frac{k(k-1)}{n}. \tag{13}$$

Proof. Any symmetric separable state is a convex combination of states of the form $\omega^n = \frac{1}{n!} \sum_\pi \tau_{\pi^{-1}(1)} \otimes \cdots \otimes \tau_{\pi^{-1}(n)}$, where $\{\tau_j\}_{j=1}^n$ is a set of pure states (these are extreme points in $\mathcal{B}(\mathbb{C}^d)$). Define $\tau := \frac{1}{n} \sum_{j=1}^n \tau_j$. We expand

$$\tau^{\otimes k} = \sum_{j_1=1}^n \cdots \sum_{j_k=1}^n M_n(j_1, \ldots, j_k) \tau_{j_1} \otimes \cdots \otimes \tau_{j_k}, \tag{14}$$

where $M_n(j_1, \ldots, j_k) = 1/n^k$ is the multinomial distribution. To compare this expression with $\omega^k := \mathsf{Tr}_{n-k}\omega^n$, write

$$\omega^k = \sum_{j_1=1}^{n} \cdots \sum_{j_k=1}^{n} H_n(j_1, \ldots, j_k)\tau_{j_1} \otimes \cdots \otimes \tau_{j_k}, \tag{15}$$

where $H_n(j_1, \ldots, j_k)$ is the hypergeometric distribution for an urn with n balls (see [1]). Then

$$\begin{aligned} \left\|\omega^k - \tau^{\otimes k}\right\|_1 &= \Big\| \sum_{j_1,\ldots,j_k} \big(H_n(j_1, \ldots, j_k) \\ &\qquad M_n(j_1, \ldots, j_k)\big)\tau_{j_1} \otimes \cdots \otimes \tau_{j_k} \Big\|_1 \\ &\leq \sum_{j_1,\ldots,j_k} \left|H_n(j_1, \ldots, j_k) - M_n(j_1, \ldots, j_k)\right| \\ &\leq \frac{k(k-1)}{n}, \end{aligned} \tag{16}$$

where we used the triangle inequality and Diaconis and Freedman's result on estimating the hypergeometric distribution with a multinomial distribution [1]. □

IV. TOWARDS A FINITE DE FINETTI THEOREM FOR THE CONVEX SETS FRAMEWORK

We will start this section with a self-contained introduction to the convex sets framework. (See Refs. [10] and [11] for a gentler introduction.) We will then generalise Lemma 5 to this setting. Finally, we pose the question of the existence of a finite de Finetti theorem in the convex sets framework.

Let Ω be the set of states of a particle. We assume that Ω is convex, compact, and has affine dimension n. In probability theory, for example, Ω is the simplex of probability distributions $(\omega_1, \ldots, \omega_{n+1})$, $\omega_i \geq 0, \sum_i \omega_i = 1$, while in quantum theory, Ω is (isomorphic to) the set of positive operators ω with trace one on a Hilbert space $\mathcal{H} \cong \mathbb{C}^d$. We are particularly interested in the case where Ω is specified by a set of conditional probability distributions $\{P_\lambda[A|X]\}$, whose elements are indexed by a label λ. This is partly because quantum states can be described in this way. For instance, the state ρ of a qubit, a spin-$\frac{1}{2}$ system, is uniquely determined by the probabilities of obtaining spin up or down when it is measured along the x, y, or z axes of the Bloch sphere. Thus a qubit can be described by a conditional probability distribution $P[A|X]$ with $\mathcal{A} = \{\uparrow, \downarrow\}$ and $\mathcal{X} = \{x, y, z\}$. Not all conditional probability distributions can be obtained by making local measurements on quantum states. This led Barrett to define generalized theories [18], where the state space Ω is the set of all conditional probability distributions $\{P_\lambda[A|X]\}$, denoted $\square$. This is the case that we considered in the previous parts of the paper. When $|\mathcal{X}| = 1$, this reduces to classical probability theory. In quantum theory, $|\mathcal{X}| = 1$ corresponds to the case where all measurements on a system commute, and thus can be performed at once. In fact, every Ω can be mapped to a convex subset of $\square$ for some number of *fiducial* measurements and outcomes [10, Lemma 1].

In quantum theory, the most general measurement that can be performed is a positive operator-valued measure (POVM), whose elements are termed *effects*. Effects are linear functions mapping states to probabilities: in (finite-dimensional) quantum theory, the probability of obtaining the outcome associated with an effect r, when the state is ω, is $r(\omega) = \mathsf{Tr}\,(R\omega)$ for some bounded nonnegative operator R with $R \leq \mathbf{1}$. In a generalized theory, effects are also functions mapping states to probabilities, and these functions should be affine so that they are compatible with preparing convex combinations. The vector space of affine functions $a : \Omega \to \mathbb{R}$, denoted $A(\Omega)$, is isomorphic to $\mathbb{R}^{n+1}$. The cone of nonnegative affine functions on Ω is denoted $A_+(\Omega)$. The *order unit* of $A(\Omega)$ is the element $e \in A(\Omega)$ satisfying $e(\omega) = 1$ for all $\omega \in \Omega$. An *effect* is an element $a \in A(\Omega)$ satisfying $0 \leq a(\omega) \leq 1$ for all $\omega \in \Omega$. The set of all effects is denoted $[0, e]$. There is a natural embedding of Ω into $A(\Omega)^*$, the dual space of $A(\Omega)$, given by $\omega \mapsto \hat{\omega}$, where $\hat{\omega}(a) = a(\omega)$ for all $a \in A(\Omega)$. Furthermore, if $\hat{\omega} \in A(\Omega)^*$ satisfies $\hat{\omega}(a) \geq 0$ for all $a \in A_+(\Omega)$ and $\hat{\omega}(e) = 1$, then $\hat{\omega}$ is the image of some state $\omega \in \Omega$ [20, Section 2.6]. We identify $\hat{\omega}$ with ω in what follows. It is easy to check that $\|\cdot\| = \sup_{a\in[0,e]} |a(\cdot)|$ is a norm on $A(\Omega)^*$. For more details about the convex sets framework, see [10, 11].

A natural distance measure on the set of states, which generalises the variational distance between classical probability distributions and the trace distance between quantum states, is given by

$$\|\omega - \omega'\| = \sup_{a\in[0,e]} |a(\omega) - a(\omega')|. \tag{17}$$

In quantum theory, systems are combined by taking the *tensor product* of the Hilbert spaces for each system. The

same is true in the convex sets framework: $\omega \otimes \omega'$ is defined to be the *product state* where system Ω is in state ω, system Ω' is in state ω', and the two systems are independent. The complication is that the space $A(\Omega)^\star$ is a Banach space but not a Hilbert space and there are multiple ways to define a norm on the tensor product space, consistent with the norm on $A(\Omega)^\star$. This choice affects the set of pure (i.e., norm 1) states of the joint system. At the very least, we want the set of joint states to be closed under convex combinations. This yields:

Definition 7. The *minimal tensor product* of Ω and Ω', denoted by $\Omega \otimes_{\min} \Omega'$ consists of all convex combinations of product states $\omega \otimes \omega'$, $\omega \in \Omega$ and $\omega' \in \Omega'$.

We say that states in $\Omega \otimes_{\min} \Omega'$ are *separable*, thereby extending terminology from quantum mechanics to the convex sets framework. Next, if a is a valid effect for system Ω and a' a valid effect for system Ω', then $a \otimes a'$ is the effect defined on product states via $a \otimes a'(\omega \otimes \omega') = a(\omega)a'(\omega')$. If all convex combinations of such effects are to be allowed, the state space must only contain states in the *maximal tensor product*, defined via duality as:

Definition 8. The *maximal tensor product* of Ω and Ω', denoted by $\Omega \otimes_{\max} \Omega'$ consists of all bilinear functions $\mu : A(\Omega) \times A(\Omega') \to \mathbb{R}$ that satisfy $\mu(a \otimes b) \geq 0$ for $a, b \geq 0$, and $\mu(e \otimes e') = 1$.

Thus $\mu \in \Omega \otimes_{\max} \Omega'$ can be written as a linear combination of product states, possibly with negative weights. In classical probability theory, the minimal and the maximal tensor product coincide. In general, a tensor product $\Omega \otimes \Omega'$ is a convex set with $\Omega \otimes_{\min} \Omega' \subseteq \Omega \otimes \Omega' \subseteq \Omega \otimes_{\max} \Omega'$. In quantum theory, $\Omega \otimes \Omega'$ is the set of trace one positive operators on the (unique) Hilbert space tensor product of $\mathcal{H}$ and $\mathcal{H}'$. Note that $\Omega \otimes \Omega'$ lies strictly between the maximal and minimal tensor products in the quantum case. The set of separable quantum states is $\Omega \otimes_{\min} \Omega'$ and $\Omega \otimes_{\max} \Omega'$ is the set of trace one entanglement witnesses.

For a state $\mu \in \Omega \otimes \Omega'$, we say that $\mu_\Omega \in \Omega$, defined by $a(\mu_\Omega) = a \otimes e'(\mu)$ for all effects a, is the *partial trace* of μ with respect to Ω'. An effect on the tensor product is an element $a \in A(\Omega \otimes \Omega')$ satisfying $0 \leq a \leq e \otimes e'$. The larger the set of joint states, the smaller the set of allowed effects. This means that the distance measure that we defined in Eq. (17), when applied to states of more than one particle, depends on which tensor product we use. It is true, however, that $\|\omega - \omega'\| \leq \|\omega - \omega'\|_{\min}$, the distance measure for the minimal tensor product, since in that case the set of effects is largest. Also note that a physical theory may place additional restrictions on which effects are allowed but, even then, $\|\omega - \omega'\|$ provides an upper bound on the probability of distinguishing ω and ω'.

Theorem 9. *Let Ω be a convex set with E extreme points (E may be infinite). Suppose $\omega^n \in \Omega^{\otimes_{min} n}$ is symmetric. Then there is a measure $m(\tau)$ on states $\tau \in \Omega$ such that*

$$\left\|\omega^k - \int dm(\tau)\, \tau^{\otimes k}\right\|_{min} \leq \min\left(\frac{2kE}{n}, \frac{k(k-1)}{n}\right). \tag{18}$$

Proof. Let $\tau_1, \ldots, \tau_E$ be the extreme points of Ω. Any symmetric separable state is a convex combination of states of the form $\omega^n = \frac{1}{n!}\sum_\pi \tau_{i_{\pi^{-1}(1)}} \otimes \cdots \otimes \tau_{i_{\pi^{-1}(n)}}$, where $1 \leq i_1, \ldots, i_n \leq E$. Define $\tau := \frac{1}{n}\sum_{j=1}^n \tau_{i_j}$. We expand

$$\tau^{\otimes k} = \sum_{j_1=1}^n \cdots \sum_{j_k=1}^n M_n(i_{j_1}, \ldots, i_{j_k}) \tau_{i_{j_1}} \otimes \cdots \otimes \tau_{i_{j_k}}, \tag{19}$$

where $M_n(i_{j_1}, \ldots, i_{j_k}) = 1/n^k$ is the multinomial distribution. To compare this expression with ω^k, write

$$\omega^k = \sum_{j_1=1}^n \cdots \sum_{j_k=1}^n H_n(i_{j_1}, \ldots, i_{j_k}) \tau_{i_{j_1}} \otimes \cdots \otimes \tau_{i_{j_k}}, \tag{20}$$

where $H_n(i_{j_1}, \ldots, i_{j_k})$ is the hypergeometric distribution for an urn with n balls (see [1]). Then

$$\begin{aligned}
\left\|\omega^k - \tau^{\otimes k}\right\|_{\min} &= \Big\| \sum_{j_1,\ldots,j_k} \big(H_n(i_{j_1}, \ldots, i_{j_k}) \\
&\qquad M_n(i_{j_1}, \ldots, i_{j_k})\big) \tau_{i_{j_1}} \otimes \cdots \otimes \tau_{i_{j_k}} \Big\|_{\min} \\
&\leq \sum_{j_1,\ldots,j_k} \left|H_n(i_{j_1}, \ldots, i_{j_k}) - M_n(i_{j_1}, \ldots, i_{j_k})\right| \\
&\leq \min\left(\frac{2kE}{n}, \frac{k(k-1)}{n}\right),
\end{aligned} \tag{21}$$

where we used the triangle inequality and Diaconis and Freedman's result on estimating the hypergeometric distribution with a multinomial distribution [1]. □

One can show that $\square^{\otimes_{\max} n}$ is precisely the set of all no-signalling conditional probability distributions and that $\square^{\otimes_{\min} n}$ is the set of all separable conditional probability distributions [18, 21]. Furthermore the trace distance (Definition 2) coincides with the definition in Eq. (17). With these observations and the fact that $||\cdot|| \leq ||\cdot||_{\min}$ we see that Theorem 9 generalises Lemma 5. Unfortunately, we were not able to obtain a similar generalisation of Lemma 4 and hence of Theorem 3. We thus conclude with the question of whether a finite de Finetti theorem exists for general theories in the convex sets framework. We remark that the argument of [9] applied in this context yields an infinite de Finetti theorem for any theory in the convex sets framework (see [12] for the details).

V. ACKNOWLEDGEMENTS

This work was carried out at the same time as related work by J. Barrett and M. Leifer [12]. We thank them for discussions, and especially for explaining how to define the trace distance. We thank R. Colbeck and R. Renner for discussions, G. Mitchison for valuable comments on the manuscript, and the organizers of the FQXi workshop *Operational probabilistic theories as foils to quantum theory*, where part of this work was done. MC thanks the IQI at Caltech and CWI Amsterdam for their hospitality. This work was supported by the UK's EPSRC, Magdalene College Cambridge, NSF Grants PHY-0456720 and CCF-0524828, EU Projects SCALA (CT-015714) and QAP (CT-015848), NWO VICI project 639-023-302, and the Dutch BSIK/BRICKS project.

[1] P. Diaconis and D. Freedman, Ann. Probab. **8**, 745 (1980).
[2] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (Wiley, Chichester, 1994).
[3] B. de Finetti, Ann. Inst. H. Poincaré **7**, 1 (1937).
[4] R. König and R. Renner, J. Math. Phys. **46**, 122108 (2005), quant-ph/0410229.
[5] M. Christandl, R. König, G. Mitchison, and R. Renner, Comm. Math. Phys. **273**, 473 (2007), quant-ph/0602130.
[6] R. Renner, Ph.D. thesis, Swiss Federal Institute of Technology, Zürich (2005), quant-ph/0512258; Nature Physics **3**, 645 (2007), quant-ph/0703069.
[7] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, 1984), pp. 175–179; A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[8] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007), quant-ph/0702152; J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005), quant-ph/0405101; Ll. Masanes, R. Renner, A. Winter, J. Barrett and M. Christandl, quant-ph/0606049; A. Acín, N. Gisin, and Ll. Masanes, Phys. Rev. Lett. **97**, 120405 (2006), quant-ph/0510094; V. Scarani, N. Gisin, N. Brunner, Ll. Masanes, S. Pino, and A. Acín, Phys. Rev. A **74**, 042339 (2006), quant-ph/0606197.
[9] C. M. Caves, C. A. Fuchs, and R. Schack, J. Math. Phys. **43**, 4537 (2002), quant-ph/0104088.
[10] H. Barnum, J. Barrett, M. Leifer, and A. Wilce (2007), quant-ph/0611295.
[11] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, Phys. Rev. Lett. **99**, 240501 (2007), arXiv:0707.0620.
[12] J. Barrett and M. Leifer (2007), arXiv:0712.2265.
[13] E. Størmer, J. Funct. Anal. **3**, 48 (1969); R. L. Hudson and G. R. Moody, Z. Wahrschein. verw. Geb. (Probab. Theory Related Fields) **33**, 343 (1976).
[14] C. A. Fuchs, R. Schack, and P. F. Scudo, Phys. Rev. A **69**, 062305 (2004), quant-ph/0307198.
[15] M. Christandl and B. Toner (2009), in preparation.
[16] T. Eggeling and R. Werner, Phys. Rev. Lett. **89**, 097905 (2002).
[17] P. Hayden, D. Leung, and G. Smith, Phys. Rev. A **71**, 062339 (2005).
[18] J. Barrett, Phys. Rev. A **75**, 032304 (2007), quant-ph/0508211.
[19] R. F. Werner, Lett. Math. Phys. **17**, 359 (1989); B. M. Terhal, A. C. Doherty, and D. Schwab, Phys. Rev. Lett. **90**, 157903 (2003), quant-ph/0210053; B. F. Toner, Proc. R. Soc. A **465**, 59–69 (2009), quant-ph/0601172.
[20] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004), available online at http://www.stanford.edu/~boyd/cvxbook/.
[21] C. H. Randall and D. J. Foulis, in *Interpretations and Foundations of Quantum Mechanics*, edited by H. Neumann (Bibliographisches Institut, Wissenschaftsverlag, Manheim, 1981).
[22] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[23] Diaconis and Freedman also obtain a second bound $k|\mathcal{A}|/n$. The analogous bound within our framework is $k|\mathcal{A}|^{|\mathcal{X}|}/n$. Restricting to *adaptive* measurements on individual particles, we are able to improve this bound to $k|\mathcal{X}|^2|\mathcal{A}|(1 + 4\sqrt{\frac{2+\log|\mathcal{X}|}{k}})/n$ [15].
[24] A density operator ρ^n is permutation-invariant if $\rho^n = \pi\rho^n\pi^{-1}$ for all permutations $\pi \in S_n$.

[25] In quantum theory, the most general measurement is termed a positive operator-valued measure (POVM). If we perform a POVM x with effects $E_{x,a}$ (satisfying $E_{x,a} = E_{x,a}^\dagger$, $E_{x,a} \succeq 0$ and $\sum_a E_{x,a} = \mathbb{1}$) on a system in state ρ, then the distribution of the measurement outcome A is given by $P[A = a|X = x] = \mathsf{Tr}\,(\rho E_{x,a})$.

[26] Let $\mathcal{A} = \mathcal{X} = \{0, 1\}$ and define $P[A_1A_2 = a_1a_2|X_1X_2 = x_1x_2] = \frac{1}{2}\,[a_1 + a_2 = x_1x_2 \pmod 2]$. This distribution is known as a nonlocal box [22] and one can easily check that it is no-signalling. We wish to distinguish this distribution from the distribution $Q[A_1A_2|X_1X_2]$, defined by $Q[A_1A_2 = a_1a_2|X_1X_2 = x_1x_2] = \frac{1}{2}\,[a_2 = 1]$. (This is an unconditional product distribution where the first bit is random and the second bit is always one.) For every setting of x_1 and x_2, $P[A_2 = 1|X_1X_2 = x_1x_2] = 1/2$, and thus P and Q cannot be perfectly distinguished by making a measurement on both systems in parallel. But if we allow adaptive strategies, then we can distinguish P and Q perfectly. For instance, set $x_1 = 1$ and then set $x_2 = a_1$, so that we have $a_1 + a_2 = 1.a_1 \pmod 2$ and it follows that $P[A_2 = 0] = 1$. Since $Q[A_2 = 0] = 0$, we conclude that we can distinguish P and Q perfectly.

[27] This immediately implies the result for $\lfloor n/|\mathcal{X}|\rfloor$. The extension to the case $\lceil n/|\mathcal{X}|\rceil$ is more technical and can be found in [15].